

# GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages

M. Frank\*, T. Lütke<sup>1</sup> and C.-W. von der Lieth

German Cancer Research Center, Spectroscopic Department (B090), Molecular Modeling Group, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany and <sup>1</sup>Bijvoet Center for Biomolecular Research, Department of Bioorganic Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received August 16, 2006; Revised October 11, 2006; Accepted October 12, 2006

## ABSTRACT

**Conformational energy maps of the glycosidic linkages are a valuable resource to gain information about preferred conformations and flexibility of carbohydrates. Here we present GlycoMapsDB, a new database containing more than 2500 calculated conformational maps for a variety of di- to pentasaccharide fragments contained in N- and O-glycans. Oligosaccharides representing branchpoints of N-glycans are included in the set of fragments, thus the influence of neighbouring residues is reflected in the conformational maps. During refinement of new crystal structures, maps contained in GlycoMapsDB can serve as a valuable resource to check whether the torsion values of a glycosidic linkage are located in an 'allowed' region similar to the Ramachandran plot analysis for proteins. This might help to improve the structural quality of the glycan data contained in the Protein Data Bank (PDB). A link between GlycoMapsDB and the PDB has been established so that the glycosidic torsions of all glycans contained in the PDB can be retrieved and compared to calculated data. The service is available at [www.glycosciences.de/modeling/glycomapsdb/](http://www.glycosciences.de/modeling/glycomapsdb/).**

## INTRODUCTION

Ramachandran plots (1) of protein backbone torsions  $\phi/\psi$  are frequently used to validate 3D structures of proteins (2,3). The quality of a protein structure is considered to be 'good' when (preferably all) amino acids have  $\phi/\psi$  values located in 'allowed' regions of the plot. For carbohydrates the glycosidic torsions  $\phi/\psi$  ( $\omega$ ) are the main determinants of the 3D structure and it is straightforward to validate the quality of an experimentally determined carbohydrate structure in a similar way to protein structures. In contrast to proteins,

however, the 'allowed' regions on a conformational map for a given glycosidic linkage not only depend on the linked monosaccharide types, but also on the linkage type and—a feature completely different from proteins—the degree of branching of the glycan. The number of available high quality crystal structures of carbohydrates is too limited to serve as a basis to determine the 'allowed' regions for all linkage types so the accessible conformational space of carbohydrate linkages has therefore to be estimated using computational methods. Consequently, over the last 30 years a considerable effort has been put into the development of force fields that are able to predict accurately the local minima and flexibility of glycosidic torsions (4). The force field most frequently used to calculate conformational maps of glycosidic linkages is MM3 (5).

## CALCULATION OF CONFORMATIONAL MAPS

A variety of methods exist to calculate the energy of a carbohydrate as a function of the glycosidic torsions  $\phi/\psi$  (or  $\phi/\psi/\omega$  for 1–6 linkages). Traditionally systematic search methods are applied to disaccharides (6). A 'relaxed' map is obtained by systematically changing  $\phi$  and  $\psi$  in small intervals (normally 10°) and minimizing all degrees of freedom while restraining  $\phi/\psi$  using an external force. Relaxed conformational maps depend on the orientation of the exocyclic torsions in the starting conformation of the carbohydrate and therefore the calculation of an 'adiabatic' map is advisable. To generate a fully adiabatic map the calculation of 3<sup>10</sup> relaxed maps would be required for a simple disaccharide (7). In routine calculations the computational cost is reduced by taking into account only *gg* and *gt* (or *gt* and *tg* for monosaccharides that have the OH4 group in an axial orientation) conformations for the hydroxymethyl groups and a *clock*- and *anticlockwise* orientation for the hydroxyl groups. This reduces the number of relaxed maps to be calculated to eight resulting in a total of 10 368 conformations to minimize for each 'pseudo' adiabatic map. It is obvious that this approach is limited to disaccharides since the number of

\*To whom correspondence should be addressed. Tel: +49 6221 424541; Fax: +49 6221 42454; Email: [m.frank@dkfz.de](mailto:m.frank@dkfz.de)

Present address:

T. Lütke, Massachusetts Institute of Technology, Biological Engineering, 77 Massachusetts Avenue, 16-719 Cambridge, MA 0213, USA

© 2006 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

conformations to be minimized for a trisaccharide would already be  $\sim 3 \times 10^6$  (assuming the second linkage is searched in intervals of  $30^\circ$ ), far too many for a large-scale project where thousands of maps need to be calculated.

High temperature molecular dynamics simulation is a robust and efficient method to explore the accessible conformational space of carbohydrates (8). Conformational free energy maps can be derived from population analysis by applying the Boltzmann equation. This approach has several advantages compared to the systematic search methods:

- it is directly applicable to branched oligosaccharides, so that the same method can be used for disaccharides and larger oligosaccharides,
- the low energy conformational space only is explored and no computational time is wasted to calculate unrealistic high energy conformations,
- the required computational cost increases therefore only moderately with the number of atoms of the oligosaccharide,
- the data-flow can be easily optimised in such a way that the conformational maps are generated automatically and only minimal human interaction is required.

It was therefore decided that free energy maps derived from MD simulations were a good first set of conformational maps with which to populate the GlycoMaps database. An in-house library of carbohydrate fragments (up to pentasaccharides) derived from structures described in the CARBBANK (9) and built using the SWEET-II program (10) served as input. The MM3 force field as implemented in the TINKER suite (dasher.wustl.edu/tinker/) [for a comparison with the original MM3 implementation see (11)] was used to calculate the trajectories at 1000 K. The length of the MD simulation was 10 ns for disaccharides and 30 ns for larger oligosaccharides. The carbohydrate rings were restrained to a chair conformation. The Conformational Analysis Tools (CAT) software (www.md-simulations.de/CAT/) was used for data processing and analysis.

## ACCESSING AND ANALYSING CONFORMATIONAL MAPS

Conformational maps of carbohydrate linkages can be retrieved from the database by entering the disaccharide fragment in IUPAC form into the search database web interface of GlycoMapsDB (Figure 1). Wildcards are supported, which allows searching, e.g. Glcp and Glcp2NAc residues simultaneously. Maps found in the database are displayed as a list, with each entry showing a preview picture and the full structure of the carbohydrate for which the map has been calculated in extended IUPAC form (Figure 2). Difference maps can be calculated e.g. to evaluate the influence of branching on the accessible conformational space of a linkage (Figure 3). Individual maps can be explored in more detail by clicking on the preview picture. A 3D structure of the carbohydrate—generated by Sweet II (10)—can be displayed using Jmol (jmol.sourceforge.net). If experimental data for the disaccharide fragment is available in the Protein Data Bank (PDB) (12), a link is displayed, which leads to a page where  $\phi/\psi$  torsion values retrieved from PDB data [using the

**Figure 1.** The search database interface of the GlycoMapsDB. Monosaccharide residues can be selected from the pull down menus or entered directly into the appropriate fields. Wildcards are supported.

Maps 71 - 80 of 94 (click on a MapID to view map details)

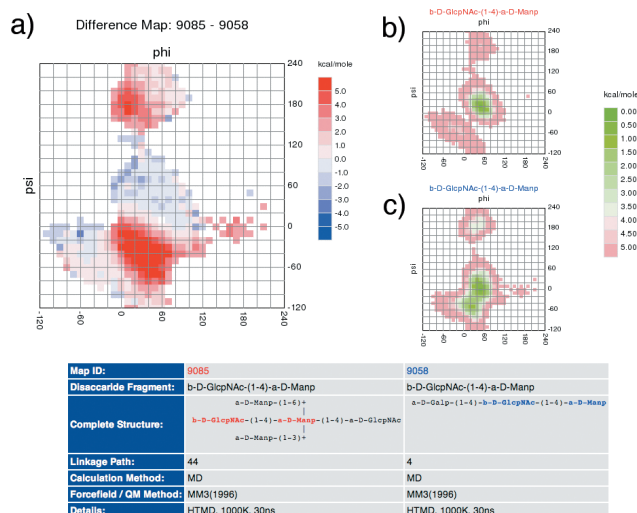
MapID	Preview	Disaccharide Fragment	Complete Structure	Method
9020		b-D-GlcpNAc-(1-4)-b-D-Manp	b-D-GlcpNAc-(1-4)+ b-D-Manp-(1-3)-b-D-Manp b-D-GlcpNAc-(1-2)+	MD MM3(1996) Tinker 4
9025		b-D-GlcpNAc-(1-4)-a-D-Manp	b-D-Glcp-(1-4)+ b-D-GlcpNAc-(1-4)-a-D-Manp a-L-Fucp-(1-3)+	MD MM3(1996) Tinker 4
9034		b-D-GlcpNAc-(1-4)-a-D-Glcp	b-D-Manp-(1-4)-b-D-GlcpNAc-(1-4)-a-D-Glcp	MD MM3(1996) Tinker 4
9058		b-D-GlcpNAc-(1-4)-a-D-Manp	a-D-Glcp-(1-4)-b-D-GlcpNAc-(1-4)-a-D-Manp	MD MM3(1996) Tinker 4
9085		b-D-GlcpNAc-(1-4)-a-D-Manp	a-D-Manp-(1-6)+ b-D-GlcpNAc-(1-4)-a-D-Manp-(1-4)-a-D-GlcpNAc a-D-Manp-(1-3)+	MD MM3(1996) Tinker 4

**Figure 2.** Example of a search result. The complete carbohydrate structure used for the calculation of the conformational map is displayed in IUPAC extended nomenclature. The disaccharide fragment that defines the glycosidic linkage is highlighted. The method used to calculate the maps is shown as well as a preview picture of the map that serves as an active link to a more detailed display of the database entry.

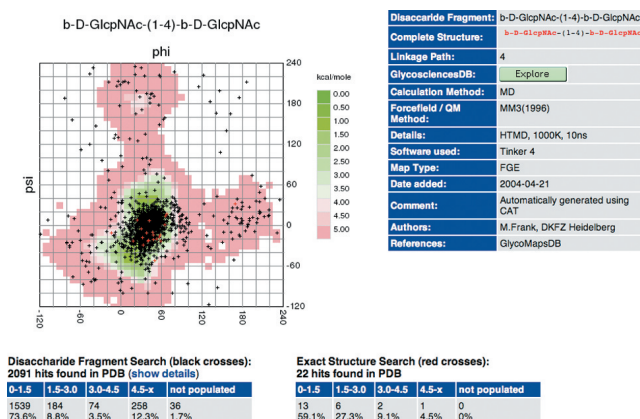
GlyTorsion tool (13)] are overlaid onto the calculated conformational map (Figure 4). Some statistical data regarding the fit of the experimental data to the calculated data is also displayed. The PDB structures evaluated can be accessed following the 'show details' link. If further information such as literature references or NMR data for the carbohydrate is available in GlycosciencesDB (14) a link to the corresponding entry in that database is displayed.

## IMPLEMENTATION

GlycoMapsDB is running on a Linux PC with Apache web server software. Interaction with the user is mediated through PHP interfaces. The datasets are stored in a MySQL database. Visualization of carbohydrate structures is performed using the java applet Jmol or the plugin Chime (www.mdlchime.com). Diagrams and plots are generated in scalable vector



**Figure 3.** Difference maps can be calculated to compare conformational maps. In the example shown the conformational maps of the disaccharide fragment  $\beta$ -D-GlcNAc-(1-4)- $\alpha$ -D-Manp either as part of a highly branched oligosaccharide (a) or as part of a linear chain (c) are compared. The reduction of accessible conformational space caused by neighbouring residues in the branched oligosaccharide can clearly be seen in the difference map (a, deep red areas represent regions where the neighbouring residue causes a strong energy penalty mainly by steric conflicts).



**Figure 4.** Comparison of calculated conformational maps with glycosidic torsion values derived from the PDB using the GlyTorsion tool (13). The PDB database entries can be explored using the 'show details' link. If available, literature references and NMR data for the carbohydrate can be retrieved from GlycosciencesDB.

graphics (SVG) format. For browsers that cannot display SVG files, they are converted to graphics interchange format (GIF) files using ImageMagick. The service is hosted at and maintained by the German Cancer Research Centre in Heidelberg, Germany. The database can be accessed online at [www.glycosciences.de/modeling/glycomapsdb/](http://www.glycosciences.de/modeling/glycomapsdb/).

## DISCUSSION AND OUTLOOK

Currently, the GlycoMapsDB contains ~2500 conformational maps of carbohydrate fragments originally described in the CARBBANK. Conformational maps for most fragments

found in glycoproteins are available in the database. The direct crosslink between calculated maps and PDB data opens a very efficient route to crosscheck the quality of experimental structures as well as the quality of the maps. In general the amount of available high quality experimental data currently available for carbohydrates in the PDB database is rather limited compared to that for proteins and for some linkages there is no experimental data available at all, so there is a clear lack of experimental reference data. In this respect, the conformational maps contained in GlycoMapsDB might help crystallographers to crosscheck their data before submission similar to the Ramachandran plot analysis for proteins. For this purpose, maps from the GlycoMapsDB are also accessed by the *carp* (Carbohydrate Ramachandran Plot) software (13), where users can upload a structure in PDB file format and retrieve plots comparing the torsions present in the structure with the conformational maps.

GlycoMapsDB indirectly offers an interface for 'data mining' in the PDB database, e.g. to find carbohydrate entries with unusual glycosidic torsion values. For  $\beta$ -D-GlcNAc-(1-4)- $\beta$ -D-GlcNAc, a frequent fragment contained in N-glycans, the agreement between experimental and calculated data is remarkably good (Figure 4). More than 80% of the crystal structures have values in low energy areas of the conformational map. But there are also some outliers that are located in 'not allowed' areas of the maps. Of special interest are also those structures that have  $\phi$  or  $\psi$  values of  $\sim 180^\circ$  (anti conformation). The 'show details' link would help to find the corresponding entries in the PDB.

In the near future we will provide also conformational maps calculated with force fields other than MM3, so that the influence of different parameter sets can be investigated. An upload feature for maps will be added so that users can upload and compare their own calculated maps with the maps stored in the database. In addition, a functionality that will allow users to calculate maps not already contained in the database is planned.

## ACKNOWLEDGEMENTS

The development of GlycoMapsDB as part of the GLYCOSCIENCES.de portal at the DKFZ was supported by a Research Grant from the German Research Foundation (DFG BIB 46 HDdkz 01-01) within the digital library program as well as the president fond of the Helmholtz society. Funding to pay the Open Access publication charges for this article was provided by the DKFZ.

*Conflict of interest statement.* None declared.

## REFERENCES

- Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) *Aspects of Protein Structure*. Academic Press, New York, pp. 121.
- Laskowski, R.A., MacArthur, M.W. and Thornton, J.M. (1998) Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.*, **8**, 631–639.
- Abola, E.E., Bairoch, A., Barker, W.C., Beck, S., Benson, D.A., Berman, H., Cameron, G., Cantor, C., Doucet, S., Hubbard, T.J.P. *et al.* (2000) Quality control in databanks for molecular biology. *Bioessays*, **22**, 1024–1034.

4. Mackerell,A.D., Jr (2004) Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.*, **25**, 1584–1604.
5. Allinger,N.L., Rahman,M. and Lii,J.H. (1990) A molecular mechanics force-field (Mm3) for alcohols and ethers. *J. Am. Chem. Soc.*, **112**, 8293–8307.
6. French,A.D. (1989) Comparisons of rigid and relaxed conformational maps for cellobiose and maltose. *Carbohydrate Res.*, **188**, 206–211.
7. Stortz,C.A. (1999) Disaccharide conformational maps: how adiabatic is an adiabatic map? *Carbohydrate Res.*, **322**, 77–86.
8. Frank,M., Bohne-Lang,A., Wetter,T. and Lieth,C.W. (2002) Rapid generation of a representative ensemble of N-glycan conformations. *In Silico Biol.*, **2**, 427–439.
9. Doubet,S., Bock,K., Smith,D., Darvill,A. and Albersheim,P. (1989) The complex carbohydrate structure database. *Trends Biochem. Sci.*, **14**, 475–477.
10. Bohne,A., Lang,E. and von der Lieth,C.W. (1999) SWEET—WWW-based rapid 3D construction of oligo- and polysaccharides. *Bioinformatics*, **15**, 767–768.
11. Stortz,C.A. (2005) Comparative performance of MM3(92) and two TINKER MM3 versions for the modeling of carbohydrates. *J. Comput. Chem.*, **26**, 471–483.
12. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
13. Lutteke,T., Frank,M. and von der Lieth,C. (2005) Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PBD. *Nucleic Acids Res.*, **33**, D242–D246.
14. Lutteke,T., Bohne-Lang,A., Loss,A., Goetz,T., Frank,M. and von der Lieth,C.W. (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology*, **16**, 71R–81R.